

## 本周工作

### 一、Map matching 调研

- 1、了解问题定义，需要考虑的方面
- 2、查到了相关比赛情况
- 3、目前方法存在的一些不足
- 4、数据的来源及形式
- 5、详情见“论文阅读部分”

### 二、邓教授访问

找到了一些科研的感觉，记了一些东西，准备写一篇博文

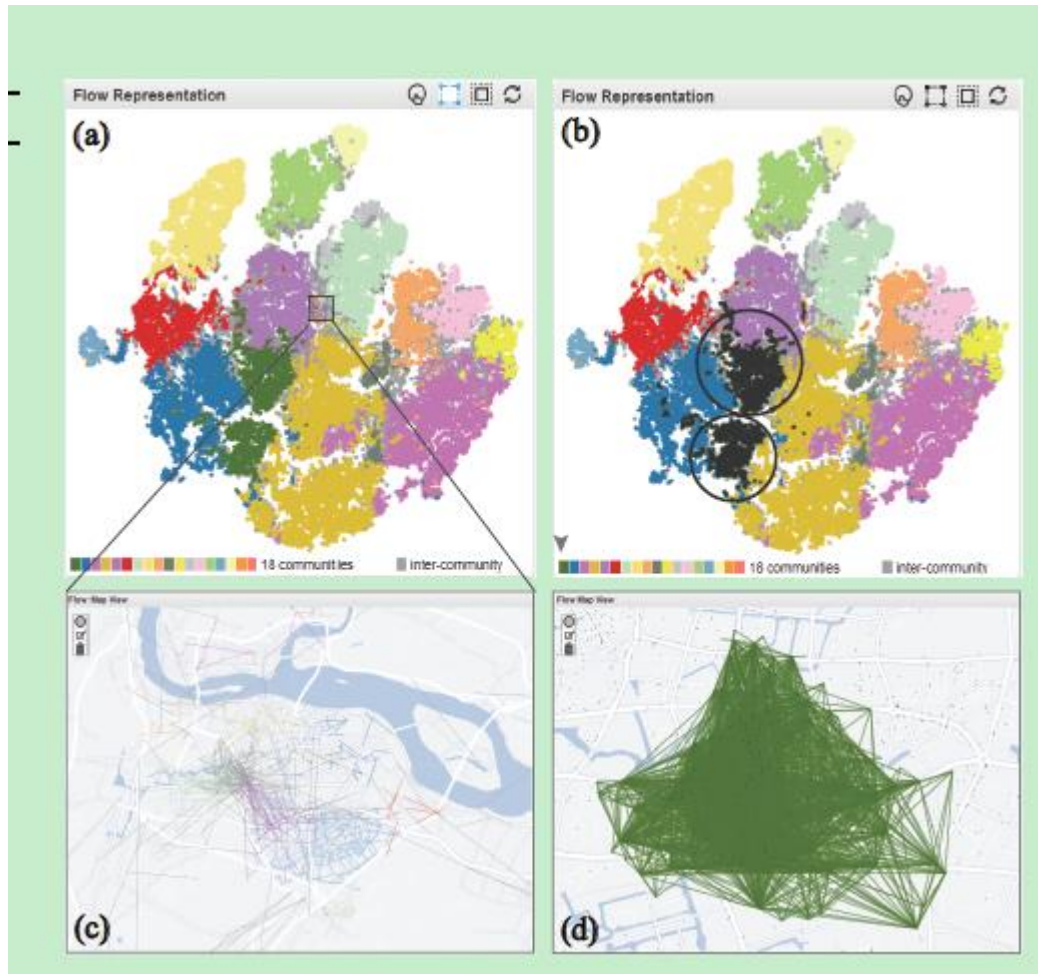
### 三、论文阅读

(1) Visual Abstraction of Large Scale Geospatial Origin-Destination Movement Data

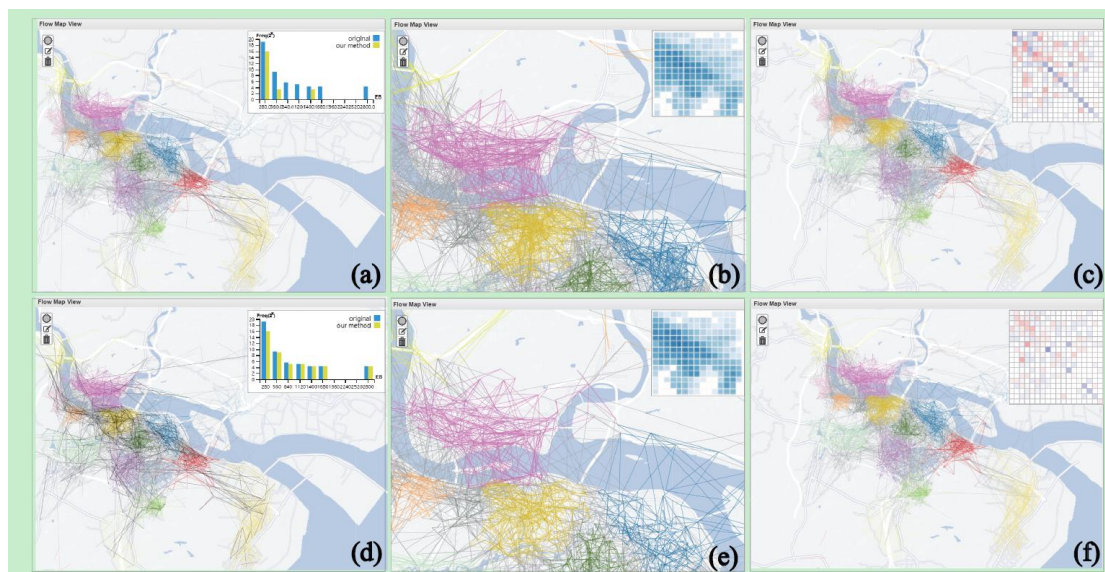
**问题：**1、许多工作在利用 OD 数据揭示模式方向都失败了

**方法：**1、word2vec, OD 数据(word), 轨迹数据(sentence)嵌入 2、蓝噪声采样法和核密度估计法相结合 3、多种约束采样约束 4、在可视化视图选择方面，选择的是较为简单的视图

**贡献：**1、将 OD 数据转换为词向量表达，有助于建立 OD 数据之间的距离关系 2、建立了距离关系则能进一步运用其他方法(引文)去分析和分类 OD 轨迹数据 4、本文为利用 OD 数据提供了迈出了有效的一步



不足：1、提出的模式对实际帮助不是特别大，比如轨迹区域的划分，如果能与演化过程相联系，则效果更佳 2、出去解释 community 以外，这样的 OD 数据不利于使用者得到有效的信息，仍显得过于杂乱 3、本图提出了多种策略去约束采样结果，以达到减少 clutter 保留相关性的目的，但针对每一种策略，本文并没有很好地描述并证明其合理性 4、可能可以借鉴思路，但本工作没有进一步提升的空间



借鉴：1、embedding 方法是一种有效的捕捉轨迹之间相互关系的方法

## 2、《A Survey of Traffic Data Visualization》

**简介：**本文是一篇交通数据可视化的 SURVEY，按照动机、交通数据预处理、交通数据可视化、交通数据可视分析的顺序，依次组织全文内容。其中对我的研究有较大帮助的章节为交通数据预处理。

本文指出，根据 2007 年的论文《Current map-matching algorithms for transport applications: State-of-the art and future research directions》，将现存的 map matching 技术分为 4 类：a)几何法 b)拓扑法 c)概率法 d)高级方法

## 3、《ACM SIGSPATIAL GIS Cup 2012、2013、2017》

**比赛简介：**为鼓励算法设计类成果，ACM SIGSPATIAL GIS 从 2012 年开始，与会议同时举行一个竞赛，赛题为 GIS 中的一些关键问题，并通过比赛推动这些问题的进一步的解决。2012 年的比赛的题目为 map matching，2013 年为 Geo-fencing、2017 年为 calculating range queries using the Fréchet distance of trajectories in mobility datasets。本次比赛同样引用了文献 2 中提到的 2007 年的综述性论文。

数据：中国出租车 GPS 数据

方法现状：没有一个方法占据支配地位

比赛定义：map matching 的基本任务是给定机动车的 GPS 轨迹路径和地图数据。把每一个地图坐标对应到此时机车最可能在的道路上

判断标准：不仅要考虑是否能够将一个点映射到一个道路上，还要分析以保证算法没有将连续的点映射到多条道路中。比较匹配的点/道路对，与实际轨迹的差别。误匹配比无匹配更危险。本竞赛采用如下形式计算得分

$$\left[ \frac{(\text{sum of confidence for correct lines}) - (\text{sum of confidence for incorrect lines})}{(\text{running time})} \right]$$

挑战：a)scalability 当采样率变化时 b)GPS 数据的固有噪声，及一些特殊地形(城市峡谷)导致的离群点 c)run in real-time, d) 生成一个有意义的置信度，以支持各种应用的使用

**Map matching 比赛结果：**

Table 1: Performance results for top five entries comparing runtime in milli-seconds, % accuracy and overall score					
Competitor	Map Load Time*	10 files, 1s Sampling**	10 files, 5s Sampling**	10 files, 10s Sampling**	100 files, 1s Sampling**
#1 - "An Efficient Algorithm for Mapping Vehicle Trajectories onto Road Networks"	(2592, 4)	(1646, 97.45%, 8.32)	(1259, 94.25%, 2.03)	(1339, 95.16%, 0.97)	(2213, 97.45%, 61.90)
#2 - "Quick Map Matching Using Multi-Core CPU"	(1683, 2)	(2080, 98.79%, 6.77)	(1823, 98.75%, 1.54)	(1885, 98.54%, 0.74)	(5135, 98.79%, 27.43)
#3 - "Effective Map-matching on the Most Simplified Road Network"	(1881, 3)	(2053, 95.54%, 6.40)	(1911, 84.53%, 1.04)	(1983, 87.83%, 0.55)	(4255, 95.54%, 30.90)
#4 - "Fast Viterbi Map Matching with Tunable Weight Functions"	(8282, 8)	(6912, 98.37%, 2.02)	(5174, 98.33%, 0.54)	(5447, 98.89%, 0.25)	(14563, 98.37%, 9.59)
#5 - "Concurrent Topological Map Matching"	(17231, 11)	(18877, 98.28%, 0.73)	(14384, 95.74%, 0.18)	(14279, 97.37%, 0.09)	(39095, 98.28%, 3.56)
* (runtime (ms), rank)		** (runtime (ms), % correct, score)			

1. 1 file, containing only 25 lines

2. 10 runs with 10 files (files for each run used a different sampling rate between 1 and 10 seconds)

3. 100 files with 1 second sampling

competition infrastructure from the ground up, so various teams could submit solutions, and developed a robust test harness that can grade various map matching solutions to the ground truth and evaluate performance along various dimensions such as correctness, speed and sampling gap. We envisioned the

比赛的重要问题：有许多参赛者将经历花在优化 IO 和查表上，虽然对时间影响大，但不是问题的核心

## 4、《Current map-matching algorithms for transport applications: State-of-the art and

future research directions》

简述：本文将 map matching 方法分为了四类, a)几何法 b)拓扑法 c)概率法 d)高级方法。并从 10 个方面指出了显存的 map matching 方法存在的问题，需要研究更加强有效的方法以满足实际的需求

#### 四、工作时间

周一至周五：9: 10-22:00      5 天 X11 小时= 55 小时

周末：14 点-21 点      2 天 X7 小时 = 14 小时

共 70 小时